Plant Physiol. (1993) 102: 387–399

Isolating the *Arabidopsis thaliana* Genes for de Novo Purine Synthesis by Suppression of *Escherichia coli* Mutants[1]

I. 5′-Phosphoribosyl-5-Aminoimidazole Synthetase

Julie F. Senecoff and Richard B. Meagher*

Department of Genetics, University of Georgia, Athens, Georgia 30602

We have initiated an investigation of the de novo purine nucleotide biosynthetic pathway in the plant *Arabidopsis thaliana*. Functional suppression of *Escherichia coli* auxotrophs allowed the direct isolation of expressed *Arabidopsis* leaf cDNAs. Using this approach we have successfully suppressed mutants in 4 of the 12 genes in this pathway. One of these cDNA clones, encoding 5′-phosphoribosyl-5-aminoimidazole (AIR) synthetase (*PUR5*) has been characterized in detail. Analysis of genomic DNA suggests that the *Arabidopsis* genome contains a single AIR synthetase gene. Analysis of the cDNA sequence and mRNA size suggests that this enzyme activity is encoded by a monofunctional polypeptide, similar to that of bacteria and unlike other eukaryotes. The *Arabidopsis* AIR synthetase contains a basic hydrophobic transit peptide consistent with transport into chloroplasts. Comparison of both the predicted amino acid and nucleotide sequence from *Arabidopsis* to those of eight other distant organisms suggests that the plant sequence is more similar to the bacterial sequences than to other eukaryotic sequences. This study provides the groundwork for future investigations into the regulation of de novo purine biosynthesis in plants. Additionally, we have demonstrated that functional suppression of bacterial mutants may provide a useful method for cloning a variety of plant genes.

The de novo pathway for purine biosynthesis provides a source of newly synthesized purine nucleotides IMP, AMP, and GMP (Fig. 1). The purine base is built up from single nitrogen and carbon units derived from one-carbon metabolism, from the amino acids Gly, Gln, Asp, and, in some cases, directly from ammonia (Neuhard and Nygaard, 1986). In spite of the increased analysis of purine biosynthesis in mammals and other eukaryotes, little is known about these pathways in higher plants (Bonner and Varner, 1976). Results of initial biochemical studies suggested that overall purine metabolism in plants is similar to that in animals, fungi, and bacteria; however, alternatives to the established steps in the pathway have been identified in diverse organisms including plants (Price and Murray, 1969; Suzuki and Takahishi, 1977; Barankiewicz and Paszkowski, 1980; Burch and Stuchbury, 1986; Moffatt and Somerville, 1988).

[1] J.F.S. was supported in part by National Institutes of Health postdoctoral fellowship GM12317. This work was supported by U.S. Department of Agriculture grant 88-37234-3548 to R.B.M.

* Corresponding author; fax 1–706–542–3910.

Purine metabolism is a central component of the nitrogen assimilation pathway in many plants. There are two characteristic pathways for nitrogen assimilation following nitrogen fixation. The first results in the production of amide compounds and the second of ureides. These compounds are then transported throughout the plant as carbon and nitrogen sources. As reviewed by Atkins (1991), the initial step in ammonia assimilation in both types of legume nodules is reductive amination to form glutamate. Because all biologically fixed nitrogen in ureide-exporting legumes (tropical legumes) must then pass through the de novo purine biosynthetic pathway, it is not surprising that the levels of purine biosynthetic enzymes are substantially increased in these nitrogen-fixing root nodules (Schubert, 1986). Purine biosynthesis in these nodules is purported to occur in the plastid of infected cells (Schubert, 1986). The end product of purine metabolism, IMP, is converted to uric acid and then to ureides, which are transported throughout the plant. By contrast, the amide exporters typically convert fixed nitrogen to Gln and Asn, which then serve as nitrogen transporters. Non-nitrogen fixing plants utilize soil ammonia and nitrates as sources of nitrogen-rich solutes for transport. For the Brassicas, including *Arabidopsis*, the primary source of transported nitrogen is Gln (C. Atkins, personal communication), which is transported to various tissues and organs of the plant through the xylem. Gln is the principal source of ammonia for de novo purine biosynthesis. As might be expected, there is a high demand for nitrogen in meristematic regions, where purine biosynthesis is required to produce DNA and RNA precursors, vitamins, and coenzymes for tissue growth and development. Presumably, these metabolic products are produced by the same or related biosynthetic enzymes that are activated during nitrogen fixation, although it is not clear whether all de novo purine biosynthesis occurs in the plastid. For a review, see Schubert (1986).

Genetic and biochemical control of de novo purine metabolism in other eukaryotes and prokaryotes is very complex

Abbreviations: AIR, 5′-phosphoribosyl-5-aminoimidazole; D_n, minimum fraction of all possible nonsynonymous nucleotide changes required to convert one amino acid coding sequence into another; GAR, 5′-phosphoribosyl-1-glycinamide; Inr, initiator; IPTG, isopropylthio-β-galactoside; LB, Luria-Bertani; pK_i, isoelectric point; SAICAR, 5′-phosphoribosyl-4-(*N*-succinocarboxamide)-5-aminoimidazole; SSC, standard sodium citrate.

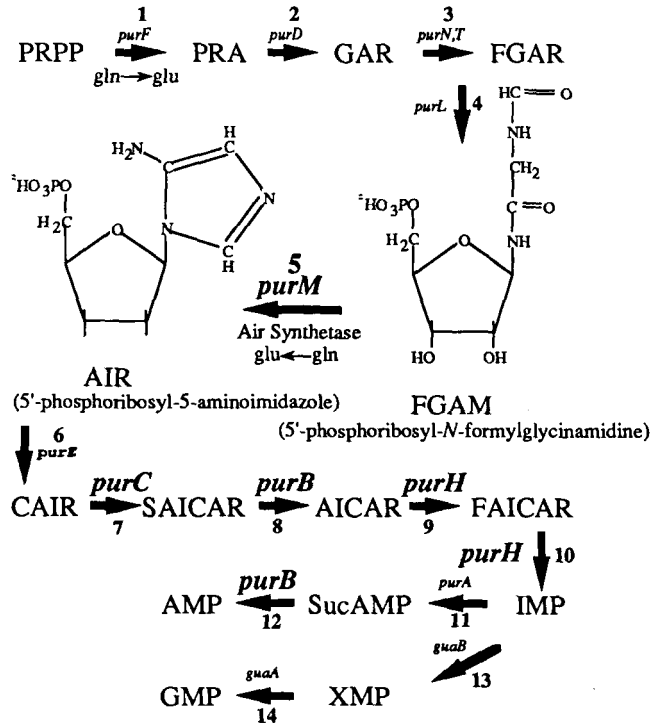

Figure 1. De novo purine nucleotide biosynthetic pathway. Genes that encode the enzymes catalyzing each step in the pathway are indicated above the reaction catalyzed. Mutant strains for the genes indicated in boldface have been successfully suppressed with *Arabidopsis* cDNA clones. The substrate and product for the reaction catalyzed by AIR synthetase (*purM*, or *PUR5*, gene) are depicted. Steps in the pathway that utilize Gln as ammonia donor are indicated. The various intermediates are abbreviated as follows: PRPP, 5-phosphoribosyl-α-1-pyrophosphate; PRA, 5-phosphoribosylamine; GAR, 5′-phosphoribosyl-1-glycinamide; FGAR, 5′-phosphoribosyl-*N*-formylglycinamide; FGAM, 5′-phosphoribosyl-*N*-formylglycinamidine; AIR, 5′-phosphoribosyl-5-aminoimidazole; CAIR, 5′-phosphoribosyl-5-aminoimidazole carboxylic acid; SAICAR; 5′-phosphoribosyl-4-(*N*-succinocarboxamide)-5-aminoimidazole; AICAR, 5′-phosphoribosyl-4-carboxamide-5-aminoimidazole; FAICAR, 5′-phosphoribosyl-4-carboxamide-5-formamidoimidazole. The abbreviations for intermediate compounds in the pathway and the gene designations are from Neuhard and Nygaard (1986). The number above each gene designation is the putative designation for the corresponding *Arabidopsis PUR* gene.

and provides many models for distinct forms of regulation (Kornberg, 1980, 1982). Control of transcription, end-product inhibition of enzyme activity, and allosteric control of enzyme activity all contribute to the regulation of purine nucleotide biosynthesis. Some of the enzymes function only when part of enzyme complexes, whereas others exist as multifunctional polypeptides. Some substrates such as 5-phosphoribosylamine are unstable (half-life of 38 s at pH 7.5, 37°C), and such intermediates may be channeled through higher-order enzyme complexes to undergo subsequent reactions (Schendel et al., 1988). Given that purines are involved in many diverse functions in plants, many control mechanisms may exist for this pathway in plants.

The availability of auxotrophic mutations in bacteria and yeast has facilitated the isolation and characterization of the DNA sequences encoding functional enzymes for de novo purine biosynthetic pathways from chicken (Chen et al., 1990) and humans (Schild et al., 1990). In this manuscript, we describe efforts to characterize the pathway for de novo purine nucleotide biosynthesis in *Arabidopsis thaliana*. We have used *Escherichia coli* mutants blocked in various steps of de novo purine biosynthesis as recipients in suppression experiments with an *A. thaliana* cDNA library to isolate functional plant cDNA sequences. *A. thaliana* is ideal for the initial investigation of this large and complex pathway for many reasons but principally because of its small genome size and thus the likelihood that many of these genes will be single copy or members of small gene families (Meyerowitz, 1987).

We report here the isolation and characterization of an *Arabidopsis* cDNA suppressing a mutation in the de novo purine biosynthetic pathway of *E. coli*. This cDNA encodes an AIR synthetase, which corresponds to the product of the *purM* gene of *E. coli* and catalyzes the closure of the large ring of the purine base (Fig. 1). The deduced amino acid and nucleotide sequences for the plant AIR synthetase are compared to those of AIR synthetase from a number of diverse organisms. This enzyme activity is part of a trifunctional enzyme in humans, chicken, and *Drosophila* and a bifunctional enzyme in yeast (Henikoff, 1986; Henikoff et al., 1986; Aimi et al., 1990; Chen et al., 1990). The plant cDNA appears to encode a monofunctional enzyme and is more similar to the bacterial enzyme (Smith and Daum, 1986; Ebbole and Zalkin, 1987). A gene tree constructed with these sequences suggests that the plant enzyme is intermediate between those encoding the bacterial enzymes and those encoding the other eukaryotic enzymes.

MATERIALS AND METHODS

Bacterial Strains and Growth Conditions

Escherichia coli strain TX709 (He et al., 1990) and other strains blocked in various steps of purine biosynthesis (Wolfe and Smith, 1985; Tiedeman et al., 1990) were the generous gift of Dr. John M. Smith (R&D Systems, Minneapolis, MN). None of the strains used showed measurable reversion frequencies of the purine auxotrophy under the conditions used ($<1/10^8$ cells). These strains were used as recipients in suppression experiments. Cells were grown in LB broth. Competent cells were prepared according the method of Dower et al. (1988). Inclusion of 1 mM Hepes (pH 7.0) buffer during the cell washes greatly increased the number of transformants obtained. Cells were frozen in 10% glycerol on dry ice and used for many months without significant loss of transformation efficiency.

cDNA and Genomic Libraries

For cDNA library preparation, RNA was isolated from 34 g of total above ground *Arabidopsis thaliana* (cv Colombia) tissue (leaves and stems) by the method of Hall et al. (1978). Poly(A)$^+$ RNA was isolated by passage over oligo(dT)-cellulose (Collaborative Research) as described by Sambrook et al. (1989). The library was prepared in the vector pcDNAII

(Invitrogen, Inc.) from 5 μg of poly(A)$^+$ RNA according to the method of Aruffo and Seed (1987). The library contained 1.7×10^7 original cDNA recombinants and was used for all suppression experiments. Plasmid DNA was prepared by growing the library in 2 L of LB containing 100 μg mL^{-1} of ampicillin overnight, followed by plasmid isolation using alkaline lysis and CsCl/ethidium bromide gradients.

A genomic DNA library was prepared in GEM11 phage vector by Dr. John Mulligan (Stanford University). The library was prepared from total *A. thaliana* (cv Colombia) DNA size selected for DNA of 15 kb or greater. After reaction with T_4 polymerase, the DNA was ligated to *Eco*RI linkers, digested with *Eco*RI, and ligated to GEM11 arms. The DNA was packaged and used to infect *E. coli* cells. The library contains approximately 200,000 original recombinants, which should represent 3×10^9 bp. Based on an accepted genome size of about 1×10^8 bp for *Arabidopsis*, this library represents about 20 to 30 genome equivalents.

Suppression of *purM* Mutations

For each transformation, 100 μL of fresh or frozen competent cells were mixed with 1 μg of plasmid DNA from the pcDNAII cDNA library in an electroporation cuvette (Bio-Rad). Electroporation was performed using the Bio-Rad Gene Pulsar set at 12.5 kV cm^{-1}, a capacitance of 25 microfaradays, and a resistance of 400 Ω. After electroporation, 1 mL of SOC broth (Sambrook et al., 1989) was added directly to the cuvette and the cells were transferred to a sterile test tube and incubated at 37°C with shaking (250 rpm) for 60 min. Inclusion of 1 mM IPTG with the SOC did not have any observable effect on the frequency of suppressing cDNAs obtained.

To identify clones that contain cDNAs suppressing the particular bacterial mutations, 100 μL of the cell suspension were plated onto each of 11 M9 (Sambrook et al., 1989) plates containing 0.4% casamino acids and 100 μg mL^{-1} of ampicillin. Again, the inclusion of IPTG (at 1 mM) in the plates had no obvious effect on transformation efficiency or suppression. Additionally, to compensate for possible catabolite repression of the *lac* promoter, both Glc and mannitol at 0.2% have been used as carbon sources. No obvious difference in transformation efficiency or suppression ability was observed between these two carbon sources. Approximately 10^5 to 10^6 transformants were spread on each plate. Appropriate dilutions were plated on LB plates containing 100 μg mL^{-1} of ampicillin to calculate the total number of transformants obtained in each experiment. Routinely, we have been able to obtain between 5×10^6 and 1×10^9 transformants from 100 μL of cells (depending on the strain used) transformed with 1 μg of purified plasmid DNA. For the second-round screening, 1 μL of miniprep DNA was used to transform 50 μL of competent cells under the same conditions as described above. Only transformants that survived the second round of selection were used in further analysis.

Sequence Analysis of *Arabidopsis PUR5* cDNA Clone

Sequencing was carried out following protocols supplied with Sequenase 2.0 (United States Biochemical). Reactions were performed using [^{32}P]dATP or [^{35}S]dATP. Products were analyzed on 6% polyacrylamide, 8 M urea gels. Gels were exposed using Kodak XAR film. Sequence alignment analyses were performed using the University of Wisconsin Genetics Computer Group sequence analysis software package (Devereaux et al., 1984).

Southern Analysis

Southern blots were performed essentially as described by Sambrook et al. (1989). Nylon Biotrans (ICN) membranes were prehybridized at 56°C in 6× SSC, 5× Denhardt's solution, 0.25% SDS, and 100 μg mL^{-1} of sheared fish sperm DNA for 1 to 12 h. Hybridization was performed under identical conditions with 2.8×10^7 cpm of radioactive cDNA insert probe. Following hybridization, the blots were washed four times for 10 min with 2× SSC, 0.5% SDS at 56°C. The blots were exposed to x-ray film for 1 to 3 d at −70°C with an intensifying screen.

RNA Analysis

RNA was extracted as described by Hall et al. (1978). Total RNA (10 μg) isolated from 29-d-old *Arabidopsis* plants and 8-d-old soybean seedlings was fractionated by electrophoresis through 1.5% agarose, 2.2 M formaldehyde gels. Hybridization was performed at 56°C in 50% formamide, 6× SSC, 5× Denhardt's solution, 0.25% SDS, 100 μg mL^{-1} of tRNA for 16 to 24 h using 10^6 cpm of probe per mL of hybridization buffer. Blots were washed four times for 10 min in 2× SSC, 0.5% SDS, 56°C, and exposed overnight to x-ray film at −70°C with an intensifying screen.

For polyacrylamide gel analysis, total RNA from each species was separated by electrophoresis in 6% polyacrylamide, 8 M urea gels. RNA was transferred to nylon membranes using a Bio-Rad electroblotter in buffer (10 mM Tris base, 5 mM acetic acid, 0.5 mM EDTA) for 3 h at 1 A. The membranes were then processed as for agarose gel northern blots.

The precise size of the *Arabidopsis PUR5* mRNA was determined using primer extension analysis (Ausubel et al., 1987). Poly(A)$^+$ (5 μg) or total RNA (50 μg) was analyzed with a primer that hybridized to nucleotides 120 to 139 of the cDNA clone. Approximately 10^6 cpm of radioactive primer were used in each annealing reaction. Products were analyzed on 6% polyacrylamide, 8 M urea sequencing gels. A sequencing ladder was used to precisely measure the size of the mRNA. Autoradiography was performed for 19 h or 3 d at −70°C with an intensifying screen.

Probes

cDNA insert probe from pcDNA709 was prepared by polymerase chain reaction amplification using oligonucleotides homologous to the T7 and SP6 promoters, which flank the cDNA inserts. Amplified fragments were purified using GeneClean (BIO101). Radioactive probes were generated using the random-primer method (Sambrook et al., 1989). Unincorporated radioactivity was removed by passing the reaction over 3-mL Sephadex G-50 columns. Oligonucleotide

probes were labeled using [γ-^{32}P]ATP and T4 polynucleotide kinase. Probes were separated from unincorporated radioactivity by chromatography on 12-mL Sephadex G-50 columns.

Chemicals and Reagents

Radionuclides were purchased from New England Nuclear. Biotrans membranes and urea were from ICN. Sequencing reagents and Sequenase were from United States Biochemical. Nitrocellulose was from Schleicher & Schuell. Chemicals, unless noted in the text, were from Sigma or Fisher. Restriction enzymes, Taq polymerase, and reverse transcriptase were from Promega or Boehringer Mannheim Biochemicals.

Quantitative Divergence among AIR Synthetase Sequences

The *Arabidopsis PUR5* protein sequence was compared to eight other sequences obtained from the National Institutes of Health and European Molecular Biology Laboratory DNA sequence data bases. The University of Wisconsin Genetics Computer Group (Devereaux et al., 1984) programs Pileup and Boxshade were used to generate aligned output of the nine sequences. A maximum of 295 amino acids among the nine sequences, starting with amino acid 66 of the *Arabidopsis* sequence, showed enough similarity to be aligned unambiguously. A few insertion and/or deletion differences separate most of the sequences, particularly the *Arabidopsis* AIR synthetase sequence, in which this region is shorter than in any of the other sequences, containing only 273 amino acids. The longest AIR synthetase sequence, *Sce*, contains 292 amino acids in the region of similarity.

Based on the amino acid alignment, the 295 codon positions or 885 nucleotides were aligned for a more quantitative assessment of sequence relatedness (Li et al., 1985) than was provided with the University of Wisconsin Genetics Computer Group package. The DNA sequences of these nine regions required some editing, because this method of DNA sequence analysis proceeds codon by codon and must be carried out on sequences of the same length. Thus, for example, 22 place-marker, XXX, codons had to be inserted into the *Arabidopsis* sequence to align it with the longer sequences from bacteria, fungi, and animals. In contrast, only three place-marker codons had to be inserted into the *Sce* DNA sequence. No codons were deleted from any of the sequences. Deletion-insertion differences between a pair of species involving a place-marker codon were not scored in any assessment of DNA divergence (i.e. these nucleotide positions were removed from both the denominator and numerator of the calculation). For example, only the 273 codons of *Arabidopsis* sequence were used in calculating actual or potential differences in sequence.

Codon-by-codon comparisons allow synonymous (silent) positions to be separated from nonsynonymous (replacement) positions, where changes cause amino acid substitutions. Therefore, divergence of the more highly conserved nonsynonymous nucleotides can be calculated free of background generated from the rapidly accumulating synonymous changes (Wolfe and Smith, 1988; Meagher et al., 1989). Considering the distance between most of the species examined, we can expect the synonymous positions to be saturated with changes (i.e. they have changed more than once). When the observed fraction of nonsynonymous changes among sequences was large, the values are adjusted for multiple hit kinetics, giving slightly higher D_n values (Li et al., 1985). A computer program generously supplied by Dr. Wen-Hsiung Li, University of Houston, was adapted to our VMS VAX system by E. McCarthy, University of Georgia.

A phenogram, or gene tree, showing the relationship among nine AIR synthetase sequences, is constructed based on the D_n values using the neighbor-joining method of Saitou and Nei (1987). The neighbor-joining method begins with a star-like tree topology among a few sequences. It constructs a binary tree by connecting sequences or groups of sequences, which successively minimize the sums of branch lengths of the intermediately produced trees. Neighbor joining is less sensitive than many other tree-building programs to small variations in evolutionary rates in different lineages, which may be important in a comparison of AIR synthetase sequences. Thus, observed differences in branch lengths are often significant. The trees produced are unrooted and appear like stars. We have drawn the AIR synthetase tree in a bush-like pattern to aid in visual interpretation. The neighbor-joining calculations were performed with a computer program kindly supplied by Masatoshi Nei and adapted to the University of Georgia VMS-VAX system by E. McCarthy and J. Arnold.

RESULTS

Isolation of cDNAs Encoding Enzymes of de Novo Purine Nucleotide Biosynthesis in *Arabidopsis* by Functional Suppression of *E. coli* Auxotrophs

A simplified pathway for the de novo biosynthesis of purine nucleotides in bacteria is shown in Figure 1. Strains of *E. coli* containing mutations in each of the 12 steps of the de novo biosynthetic pathway were used as recipients in suppression experiments. Electroporation-competent cells were prepared (Dower et al., 1988), and plasmid DNA from a cDNA library containing 1.7×10^7 original *Arabidopsis* leaf cDNA clones was used to transform each mutant strain. Approximately 10^7 to 10^8 total transformants were plated on M9 plates with 100 μg mL^{-1} of ampicillin lacking exogenous purine sources. The plates were incubated at 37°C for 16 to 96 h. Colonies that appeared on minimal-medium plates were streaked onto a second minimal-medium plate. Those cells that grew a second time were characterized further, and plasmid DNA was isolated from these cells. Small amounts of plasmid DNA were used to transform the same recipient strain, again selecting for suppression of the purine auxotrophy. In this second round of transformation, only clones that gave a high frequency of suppression were selected and subjected to further analysis.

Suppressing clones were obtained at frequencies varying 1 in 10^5 to 1 in 10^6 total transformants (Table I). We have isolated plasmids containing *Arabidopsis* cDNAs that suppress the auxotrophs *purM, purB, purC,* and *purH* of *E. coli*. All of the *E. coli* strains containing mutations suppressed by the *Arabidopsis* plasmid cDNA produced visible colonies in 16 to

Table I. *Frequency of suppressing clones obtained for four steps in purine biosynthesis*

Strain names and genotypes are as indicated. Number of transformants is the total number of ampicillin-resistant colonies obtained as determined by plating appropriate dilutions on LB plates containing 100 $\mu g\ mL^{-1}$ of ampicillin. Frequency of clones is the number of suppressing clones obtained per total number of transformants for each strain.

Strain	Genotype	Number of Transformants	Frequency of Clones
TX530	*araD (lac) purB-lacZ⁺Y⁺*::1 p1 (209)	6×10^7	$1/3.3 \times 10^5$
TX709	*supD (lac)* U169 *purM-lacZ⁺Y⁺*::Kan^r	8.9×10^6	$1/2.7 \times 10^5$
TX717	*supD (lac)* U169 *purC-lacZ⁺Y⁺*::Kan^r	1.3×10^8	$1/1.0 \times 10^5$
TX732	*araD*139 Δ(*ara-leu*) 7696 Δ (*lac*) 74 *galU galK hsr⁻ hsm⁺ mcrA⁻ mcrB⁻ StrA purH-lacZ⁺Y⁺*::Kan^r	1.9×10^8	$1/1.0 \times 10^6$

24 h at 37°C on minimal-medium plates. Original transformants that took longer to grow on minimal-medium plates after the initial transformation did not contain plasmid DNA. A plasmid designated pcDNA709, encoding a presumptive plant AIR synthetase (*PUR5*), was examined in greater detail and is the subject of the remainder of this paper.

Sequencing and Characterization of cDNA Encoding AIR Synthetase

Figure 2A presents the structure of the 1.5-kb cDNA insert in pcDNA709, which gives high-frequency transformation of TX709 (*purM* mutant) on minimal-medium plates. It is interesting that the orientation of this cDNA is opposite to the direction of transcription from the β-galactosidase promoter. This indicated that synthesis of the *Arabidopsis* AIR synthetase protein in *E. coli* occurred through transcription and translation signals independently of those associated with the β-galactosidase signals provided in this vector. A similar phenomenon was reported in the cloning of human de novo purine synthesis cDNAs (Aimi et al., 1990). We do not know whether all of the clones are in this reverse orientation, but clearly sufficient expression of this gene occurs to suppress the null mutation of this *E. coli* strain.

One plasmid clone isolated that suppressed an *E. coli purM* mutation contained 1467 nucleotides (Fig. 2B), including 15 adenosine residues at the 3′ end of the clone, corresponding to the poly(A) tail of the mRNA. This clone has been designated pcDNA709. The cDNA sequence contained 134 bp of upstream sequence before the first ATG occurred. This first ATG codon began an open reading frame that would potentially encode a 355-amino acid protein. On the basis of the evidence obtained from suppression and sequence analysis, we inferred that this open reading frame encoded an *Arabidopsis PUR5* gene. Following conventions for nomenclature in *Arabidopsis* we have designated this gene *PUR5*, because this gene encodes the fifth step of the purine biosynthetic pathway as defined in *E. coli*.

Isolation and Characterization of AIR Synthetase Genomic Clones

To verify the sequence of the AIR synthetase cDNA clone and to analyze the regulatory sequences, the entire sequence of an AIR synthetase genomic clone from *Arabidopsis* was also obtained. The coding portions of this genomic sequence were in complete agreement with the sequence of the cDNA clone. The genomic clone was shown to have seven introns and eight exons (Fig. 2A). The introns are all small, containing 87, 72, 83, 72, 73, 82, and 83 bp for intron 1 through intron 7. Introns 1 and 5 both split codons, intron 1 a Gly codon, G/GT, and intron 5 an Arg codon, AG/G. Both of these intron insertions occur in the same sequence context, AG/G, where the intron is inserted after the AG. Sequence analysis has demonstrated that all of the introns follow the GT-AG sequence rule for splice junctions, and all have an A + T composition greater than 60%.

The first 58 amino acids of the presumed open reading frame contained several features of plant transit sequences (Berry-Lowe and Schmidt, 1991) required for transport of proteins encoded by nuclear genes into chloroplasts. The overall amino acid composition of this region was basic, with a calculated pK_i of 11.6. The basic nature of this polypeptide is not characteristic of the remainder of the mature AIR synthetase sequence, which had a calculated pK_i of 4.64. It is interesting that the pK_i of the *E. coli* AIR synthetase is 4.61, very similar to that for the mature *Arabidopsis* polypeptide. By contrast, the yeast enzymes have a pK_i near 5.0, and the higher eukaryotic enzymes have pK_i values near 7.0 (Table II). Of the 58 amino acids in the potential transit sequence, 20 are Ser and Thr residues. A similar high proportion of Ser and Thr residues has been observed in other transit peptides for chloroplast-directed proteins (Berry-Lowe and Schmidt, 1991). We have observed no homology in this region when compared to a variety of AIR synthetase sequences (see below). The identification of a potential chloroplast transit peptide was not unexpected, because it is likely that de novo purine synthesis and many other important reactions involving nitrogen metabolism and assimilation (Schubert, 1986) occur in the chloroplast.

The 3′ region of the cDNA contains several elements that are considered to be important determinants for directing polyadenylation in plants. Sequence analysis of the 3′ region of the cDNA clone demonstrated a GT-rich tract (40 of 46 nucleotides) positioned 70 to 115 bases downstream of the TAG stop codon. Two potential polyadenylation signals, AAGAA and AATTAA, which overlap one another, were identified starting 127 nucleotides after the termination codon. The poly(A) tract identified in the cDNA began at an A residue 83 bases downstream of the AATTAA sequence. Previous work has demonstrated that polyadenylation determinants in plants are not as highly conserved or precise as those found in animal systems (Shirley et al., 1990).

Southern Blot Analysis of *PUR5* Sequences in *Arabidopsis*

Genomic DNA samples were digested with a variety of restriction endonucleases, separated on 0.8% agarose gels, transferred to nylon membranes, and probed with a radiola-

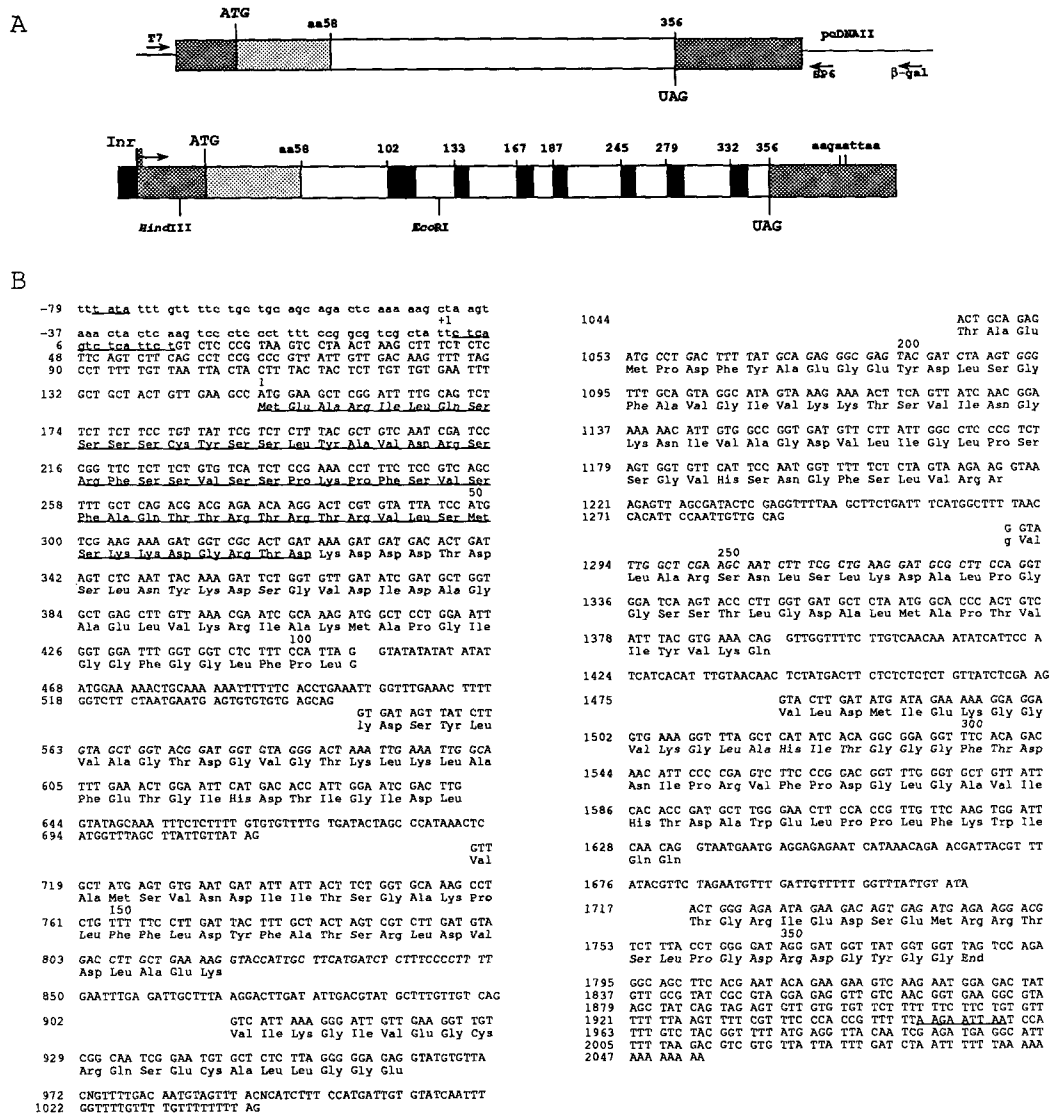

Figure 2. Structure of *Arabidopsis PUR5* cDNA and genomic DNA clones. A, Physical map of *PUR5* cDNA and genomic DNA clones. Regions encoding the putative mature AIR synthetase protein (white), the potential transit peptide sequence (light shading), and the 5′- and 3′-untranslated flanking sequences (dark shading) are shown. Intron elements are indicated by black. The amino acid position at which an intron occurs is indicated at the beginning of the intron. Elements that may be involved in promoter regulation at the 5′ end (Inr) or polyadenylation at the 3′ end are indicated (aagaattaa). The black region represents uncharacterized upstream sequences contained within the genomic clone and not part of the cDNA. B, Complete sequence of *PUR5* genomic DNA clone. The sequence and physical structure presented were obtained by analysis of both the genomic and cDNA sequences. The putative transcription start site inferred from primer extension analysis and mapped on the genomic sequence is indicated with +1. Lowercase letters at the 5′ end of the sequence are contained within the genomic clone and not the cDNA clone. Regions of interest are underlined as follows: within the 5′-untranslated region at −75 relative to the transcription start is a TATA element. Also, at +2 are two copies of the initiator sequence element (CTCANTCT). A putative transit peptide sequence comprising amino acids 1 through 58 is underlined. Sequences potentially involved in directing poly(A) addition in the 3′ region are also underlined. The amino acid numbers are indicated above the corresponding codons.

Table II. *Physical characteristics of AIR synthetase peptide sequences from nine organisms*

Only the portions of the multifunctional proteins containing the AIR synthetase domain were used in these calculations. For this comparison, the region of the *Arabidopsis* AIR synthetase lacking the putative transit peptide was used.

Organism	Residues	Mol Wt	pK_i	Charge
A. thaliana	297	31,283	4.64	−12
E. coli	345	36,821	4.65	−18
Bacillus subtilis	346	37,052	4.61	−21
Gallus gallus	372	39,359	7.21	−1
Homo sapiens	376	40,113	7.18	−1
S. cerevisiae	360	38,449	4.66	−16
Schizosaccharomyces pombe	359	38,996	5.10	−12
D. melanogaster	355	37,842	6.67	−4
D. pseudoobscura	355	37,492	6.77	−3

beled *PUR5* cDNA. A representative blot is shown in Figure 3. The enzymes *Eco*RI and *Hin*dIII both cut within the cDNA sequences, resulting in two bands that hybridize on the genomic Southern blot. The DNA digested with *Bam*HI yielded only high mol wt DNA and did not give an interpretable pattern even after longer exposures. The sizes of the *Hin*dIII and *Eco*RI digestion products were confirmed by Southern blot analysis of two independently isolated λ genomic clones containing *Arabidopsis PUR5* sequences (data not shown). Comparison with copy reconstructions using the plasmid containing the *PUR5* cDNA insert indicated that the

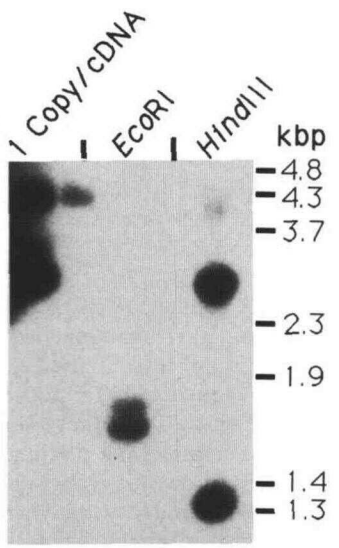

Figure 3. Genomic Southern blot. Five micrograms of total *Arabidopsis* genomic DNA was digested with restriction endonucleases, separated by electrophoresis, and blotted to filters. The blot was probed with a polymerase chain reaction product generated from pcDNA709 AIR synthetase cDNA using oligonucleotides homologous to the T7 and SP6 promoters, which flank the cDNA inserts. A single-copy reconstruction based on the haploid *Arabidopsis* genome size (1×10^8 bp) was from the plasmid pcDNA709. Molecular size standards are indicated.

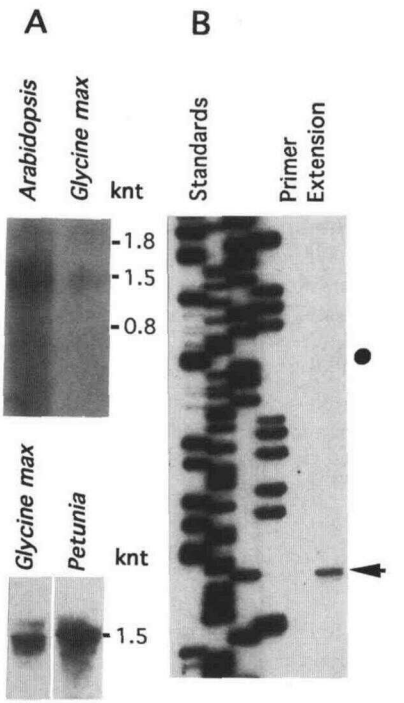

Figure 4. Analysis of *pur5* mRNA A, Northern blot of *Arabidopsis* and soybean *PUR5* RNAs. Ten micrograms of total RNA isolated from 4-week-old *A. thaliana* plants or 9-d-old *G. max* seedlings was separated on 1.5% agarose, 2.2 M formaldehyde gels. The blot was probed with the pcDNA709 insert (see Fig. 3). Molecular size standards are indicated. For polyacrylamide gel analysis, 10 μg of total RNA isolated from 9-d-old *G. max* seedlings or mature *P. hybrida* plants were separated on 6% polyacrylamide, 8 M urea gels. The RNAs were transferred by electroblotting to nylon membranes and probed as above. B, Primer extension mapping of the 5′ end of the *PUR5* mRNA. Five micrograms of poly(A)$^+$ RNA were hybridized with 10^6 cpm of end-labeled oligonucleotide that hybridized 139 bp from the 5′ end of the cDNA clone. Primer extension reactions were performed as described by Ausubel et al. (1987). The major product observed is 15 nucleotides longer than the cDNA clone. The expected initiation sequence for transcription directed by the TATA element shown in Figure 2B is indicated with a dot. See Figure 2B for the sequence corresponding to the transcription initiation site. knt, Kilonucleotides.

Arabidopsis genome probably contains a single copy of the *PUR5* sequence.

Northern Blot Analysis of *PUR5* in Total RNA from *Arabidopsis*, *Glycine max*, and *Petunia hybrida*

As stated previously, the AIR synthetase activity in other eukaryotic systems exists as part of a larger polypeptide, possessing one (yeast) or two (human, chicken, *Drosophila*) additional enzyme activities. To ascertain whether the cDNA isolated was a truncated portion of a larger mRNA, northern blot analysis was performed using agarose formaldehyde gels. The results of this experiment are shown in Figure 4A. The mRNA homologous to the *PUR5* cDNA probe migrated to a position of approximately 1.5 kb, consistent with the size

of the cDNA clone isolated. Additionally, the cDNA probe isolated from *Arabidopsis* hybridized to an mRNA of similar size from *G. max* (Fig. 4A). The hybridization data supported the hypothesis that the 1.5-kb *Arabidopsis* AIR synthetase cDNA clone, pcDNA709, was nearly full length and encoded a monofunctional polypeptide in plants. Using denaturing PAGE followed by electroblotting, we were able to detect RNA molecules homologous to the *Arabidopsis PUR5* cDNA in *G. max* and *P. hybrida* (Fig. 4A) under moderately stringent hybridization conditions. This gel system does not predict the mol wt of these transcripts as accurately as the agarose system because of the poor resolution of high mol wt molecules (greater than 1 kb). The increased sensitivity of this system provided better detection of these heterologous RNAs. However, the hybridization to *Arabidopsis* RNA was significantly stronger (data not shown). This indicates that the homology to the other plant genes may be somewhat limited.

Primer Extension Analysis of *Arabidopsis PUR5* mRNA

The cDNA isolated and characterized in pcDNA709 is of comparable size to the mRNA product discerned by northern blot analysis. However, small differences in size between the cDNA and mRNA would not be revealed by this analysis. We have utilized primer extension reactions to map the 5′ end of the mRNA precisely and determine the start of transcription.

A representative primer extension reaction is shown in Figure 4B. The principal product generated from primer extension reactions on the poly(A)$^+$ RNA fraction was 154 nucleotides in length based on the sequencing ladder. The distance from the 5′ end of this primer to the end of the cDNA sequences is 139 nucleotides. Thus, the mRNA encoding AIR synthetase in *Arabidopsis* leaf tissue was shown to be 15 nucleotides longer than the cDNA clone isolated. On the basis of the genomic sequence (see below), we have determined that, in the absence of an intron within this 5′-untranslated region, transcription initiates at the thymine nucleotide indicated in Figure 2B. Extended exposures of this gel revealed the presence of two other minor primer extension products, one of 153 nucleotides and one of 157 nucleotides. These products may reflect a small amount of heterogeneity in the transcription start site (see below) or artifacts generated by the reverse transcriptase reaction. Analogous reactions performed using total RNA samples yielded similar results (data not shown).

Genomic AIR synthetase clones were isolated, using the cDNA insert in pcDNA709 as a probe, to discern the exact sequences surrounding the 5′ end of the mRNA. Sequence analysis of the proximal promoter region revealed some unique features for the *PUR5* gene promoter. There is no discernible homology to the consensus TATA box at the conventional location near −35 relative to the transcription start site (Fig. 2B). However, two perfect matches to the 7-bp Inr sequence, CTCANTCT, were contained in the region immediately following the transcription start site (Fig. 2B). These two Inr motifs overlap each other by 2 bp. The Inr sequence element has been previously identified in mammalian promoters (Pugh and Tjian, 1990; Smale et al., 1990), and variations on this sequence have been found in *Aspergillus* (Adams and Timberlake, 1990). The Inr motif has been shown to direct transcription initiation in other TATA-less promoters as well as in promoters containing TATA boxes (O'Shea-Greenfield and Smale, 1992; Zenzie-Gregory et al., 1992).

A potential TATA sequence is present within the genomic DNA sequence at about −75 relative to the presumed transcription initiation site (and the Inr sequences, Fig. 2B). Examination of the region of the gel revealed no secondary start sites corresponding to transcription directed by this element (Fig. 4B). Lengthy exposures of this same gel also did not reveal any additional transcripts (data not shown). Although we cannot rule out the presence of an intron within this region, we believe that it is most likely that the Inr elements are responsible for directing transcription initiation from this gene.

Relationship of the *Arabidopsis* AIR Synthetase Sequence to Those from Other Kingdoms

The *Arabidopsis* (Ath) *PUR5* amino acid sequence was compared to other known AIR synthetase sequences (Fig. 5). Genes representing four animal, two fungal, and two bacterial sequences were found in the GenBank and EMBL DNA data bases. The four animal AIR synthetase sequences from *Drosophila melanogaster* (Dme) (Henikoff et al., 1986), *Drosophila pseudoobscura* (Dps) (Henikoff and Eghtedarzadeh, 1987), chicken (Gga) (Chen et al., 1990), and human (Hsa) (Schild et al., 1990) are each part of a much larger gene and transcript encoding a trifunctional enzyme with not only AIR synthetase but GAR synthetase and GAR transformylase enzyme activities (Daubner et al., 1985; Henikoff, 1986). The two fungal AIR synthetase sequences from *Schizosaccharomyces pombe* (Spo) and *Saccharomyces cerevisiae* (Sce) (Henikoff, 1986) are each part of genes encoding bifunctional enzymes. Only the portions of the fungal and animal sequences thought to encode AIR synthetase activity were included in the comparison. The two bacterial AIR synthetase genes from *E. coli* (Eco) (Smith and Daum, 1986) and *Bacillus subtilis* (Bsu) (Ebbole and Zalkin, 1987) encode AIR synthetase alone.

The *Arabidopsis* AIR synthetase sequence (excluding the predicted transit peptide) gave unambiguous alignment with the other eight protein sequences. Of the 297 amino acids in this region of the Ath sequence, 159 residues were identical with the Eco sequence and 153 residues were identical with the Hsa sequence. It was interesting that the optimum multiple alignment of these nine sequences suggested that the plant sequence was most similar to the bacterial sequences and most distinct from the animal sequences (see "Discussion"). It was also clear that the nine AIR synthetase sequences examined were homologous sequences sharing a single common evolutionary origin. AIR synthetase represents a moderately conserved polypeptide sequence, considering the small number of amino acid changes that have occurred among the four kingdoms examined (Margulis and Schwartz, 1982).

To quantify further the divergence of the nine AIR synthetase genes, the DNA sequences encoding this conserved region were compared for nonsynonymous nucleotide sub-

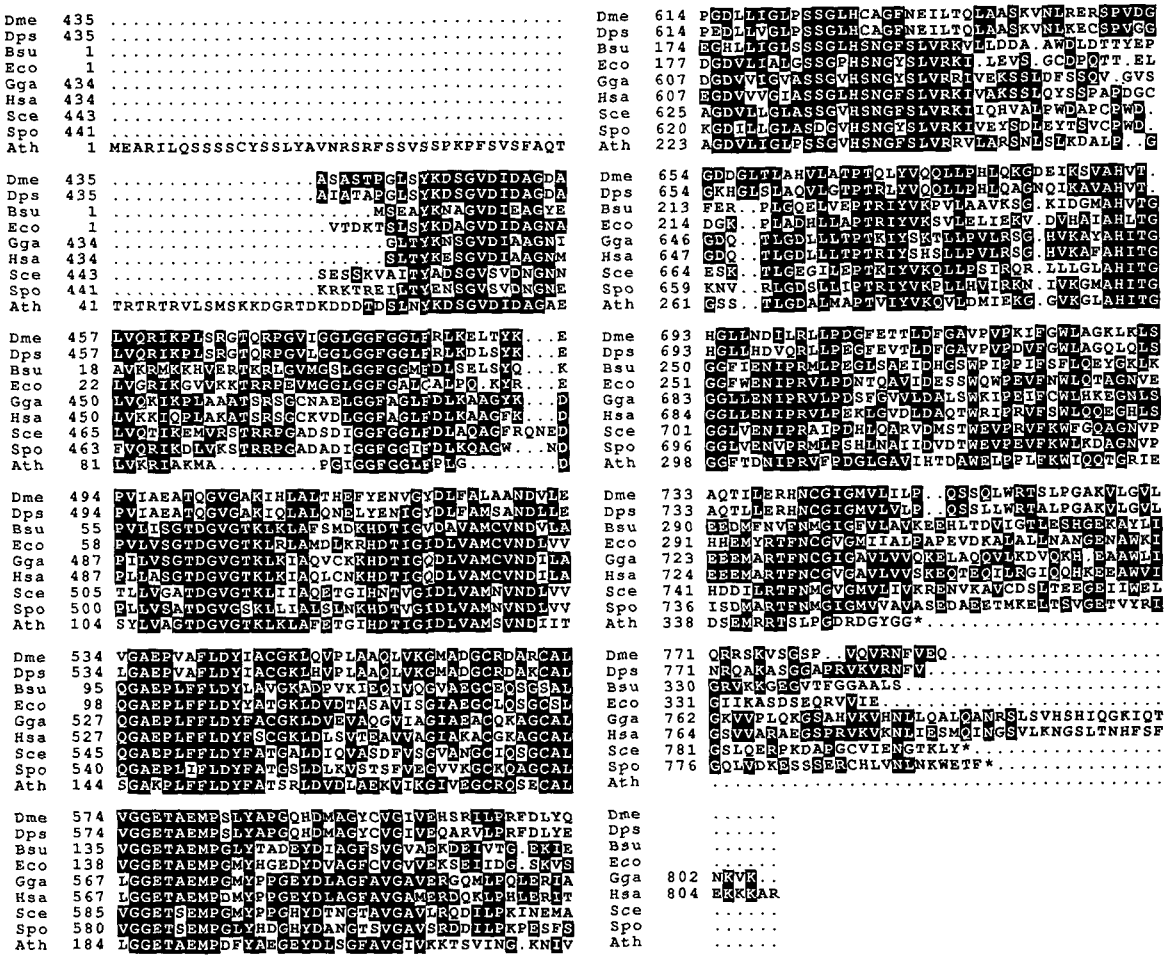

Figure 5. Sequence comparison for nine AIR synthetase enzymes. The AIR synthetase amino acid sequences from four animals (Dme, Dps, Gga, Hsa), two fungi (Sce, Spo), *Arabidopsis* (Ath), and two bacteria (Bsu, Eco) were compared (see "Materials and Methods"). The first 58 amino acids of the presumed *Arabidopsis* chloroplast transit precede a moderately conserved region of homology among all the sequences. Residues that are identical in at least two of the species are shaded. Deletions in any one sequence relative to the others are marked with periods(.). The AIR synthetase domains from the multifunctional enzymes were isolated and the sequences were aligned using Pileup. The numbers to the left of the sequence correspond to the amino acid number of the full-length protein. See text for details.

stitution differences (Li et al., 1985). This approach calculates D_n and was corrected for multiple-hit kinetics. The data are presented in Figure 6A. Even with the higher sensitivity obtained by examining D_n instead of amino acid differences, the Ath sequence was more similar to the bacterial sequences (D_n = 0.38–0.44) and most distinct from animal sequences (D_n = 0.44–0.59). The potential significance of this finding will be discussed below.

AIR Synthetase Gene Phylogeny

A phenogram based on D_n values among AIR synthetase sequences is shown in Figure 6B. This tree was constructed using the neighbor-joining tree-building method (Saitou and Nei, 1987). Although the bacterial sequences are on one extreme of the gene tree relative to animal sequences, they were barely resolved from the Ath sequence, again suggesting a close relationship among the plant and bacterial sequences. Furthermore, the length of the Ath branch relative to those for the bacterial, fungal, and most animal sequences suggests that the nonsynonymous positions in the plant sequences have evolved at normal rates for an AIR synthetase gene.

DISCUSSION

Using functional suppression of *E. coli* mutant strains deficient in purine biosynthesis, we have succeeded in cloning the gene encoding AIR synthetase from *A. thaliana*. A nearly full-length *PUR5* cDNA was isolated and shown to encode a monofunctional polypeptide for AIR synthetase. Previous work has demonstrated that other eukaryotic AIR synthetase activities are encoded as part of a larger, multifunctional enzyme (Henikoff, 1986; Henikoff et al., 1986; Aimi et al., 1990; Chen et al., 1990). By contrast, the plant gene resembles more closely the structure of the prokaryotic enzyme, where

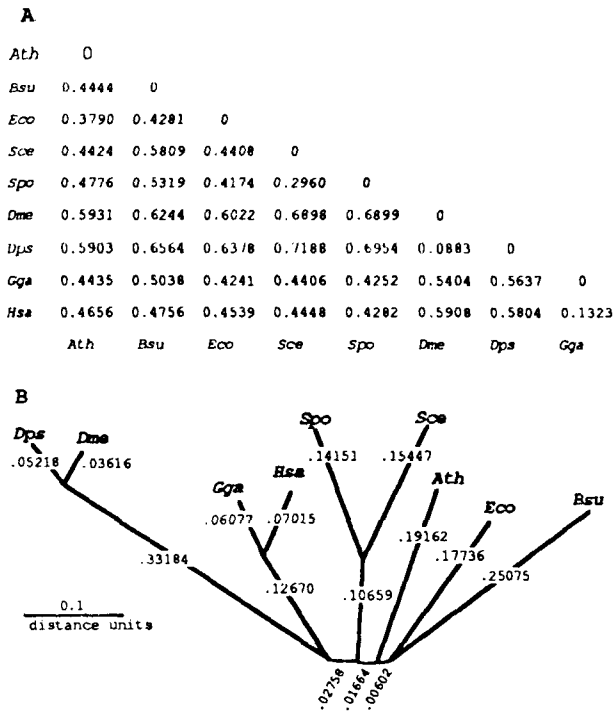

Figure 6. Quantitative comparisons of nonsynonymous nucleotide positions among AIR synthetase sequences. A, Fraction of nonsynonymous nucleotide changes (D_n) for pairwise comparisons of nine AIR synthetase DNA sequences obtained from four animals (D. melanogaster, Dme; D. pseudoobscura, Dps; G. gallus, Gga; H. sapiens, Hsa), two fungi (S. cerevisiae, Sce; S. pombe, Spo), one plant (*A. thaliana*, Ath), and two bacteria (B. subtilis, Bsu; *E. coli*, Eco) were calculated (see "Materials and Methods"). The alignments of Ath *PUR5* residues 66 to 348 shown in Figure 5 were used in this codon-by-codon analysis of 859 nucleotides of DNA sequence. The total number of codon positions compared is determined by the longest possible sequence alignment. B, A neighbor-joining gene tree showing the relationship among AIR synthetase sequences based on nonsynonymous nucleotide differences (A). Fractional differences in D_n are presented (i.e. 0.1 distance units represents approximately 10% difference in nonsynonymous nucleotide positions corrected for multiple-hit kinetics).

the gene encodes a monofunctional AIR synthetase (Smith and Daum, 1986; Ebbole and Zalkin, 1987).

We have also suppressed mutations in several other *E. coli* strains. The activities we have isolated are SAICAR synthetase (*E. coli purC* gene, strain TX717 [He et al., 1990]; *Arabidopsis PUR7* gene), adenylosuccinate lyase (*E. coli purB* gene, strain TX530 [Wolfe and Smith, 1985]; *Arabidopsis PUR8* gene), and 5'-phosphoribosyl-4-carboxamide-5-aminoimidazole transformylase IMP cyclohydrolase (*E. coli purH* gene, strain TX732 [unpublished]; *Arabidopsis PUR9* gene) (Fig. 1). A number of chicken *pur* genes have been isolated by suppression of *E. coli pur* mutations (Chen et al., 1990), whereas human and *Drosophila pur* cDNAs have been cloned by suppression of yeast *ade* mutations (Henikoff et al., 1981; Schild et al., 1990; Barton et al., 1991). In plants, suppression of *E. coli* mutant phenotypes has been used successfully to isolate the phospho*enol*pyruvate carboxylase cDNAs from a maize library (Izui et al., 1986) and a superoxide dismutase cDNAs from *Arabidopsis* and tobacco libraries (Van Camp et al., 1990). Recently, several yeast auxotrophs, including one involved in pyrimidine biosynthesis, were functionally suppressed using an *Arabidopsis* cDNA library (Minet et al., 1992). It is possible that functional suppression of mutations in other steps in the de novo purine biosynthetic pathway still can be obtained using alternate vectors or other plant cDNA libraries, perhaps constructed from organs other than leaves, where the levels of purine biosynthetic enzymes may be elevated.

Based on the analysis of cDNA and genomic clones and the identification of a single gene on Southern blotting of genomic DNA it appears that *Arabidopsis* contains a single copy of the AIR synthetase (*PUR5*) gene. The lack of additional hybridization signals with the cDNA probe indicates that any additional *PUR5* genes in the genome would have to be significantly diverged not to be detected under the hybridization conditions used in these experiments.

Arabidopsis PUR5 Gene Encodes a Monofunctional AIR Synthetase

The 1.5-kb *PUR5* cDNA isolated encoded a complete AIR synthetase enzyme and a basic amino-terminal peptide with characteristics of plant transit sequences. Several other independently isolated *Arabidopsis PUR5* cDNAs that suppressed the TX709 mutation in *purM* (not shown) were of similar size to the one characterized here. The *Arabidopsis PUR5* cDNA hybridized at moderate stringency to an RNA molecule of approximately 1.5 kb in *Arabidopsis*, *P. hybrida*, and *G. max* total RNA preparations. It was not surprising, considering the amino acid conservation among all of the AIR synthetase sequences examined across kingdoms (Fig. 5), to find this sequence conserved among these dicots. These plants are assumed to have shared a common ancestor within approximately 65 to 110 million years (Cronquist, 1981).

On the basis of these data, it is likely that this transcript represents the correct size of the mature *PUR5* mRNA in all dicots. It seems likely, taking into account the transcript size and the homology of the predicted amino acid sequence of the *Arabidopsis* protein to the monofunctional bacterial AIR synthetase sequences, that most higher plants will encode a monofunctional AIR synthetase enzyme. In other eukaryotic organisms, AIR synthetase activity is found as part of a single multifunctional polypeptide in combination with one or two other enzymic activities (Henikoff, 1987). For example, in animals the activities of GAR synthetase (*purD*, Fig. 1), GAR transformylase (*purN*), and AIR synthetase (*purM*) are part of a single multifunctional polypeptide (Daubner et al., 1985; Henikoff et al., 1986). In yeast only two enzyme activities, GAR synthetase and AIR synthetase, are part of this multifunctional protein (Henikoff, 1986). None of our data suggested that the plant *PUR5* transcript was part of a larger sequence encoding a multifunctional enzyme in plants. Additionally, we have obtained no evidence for a second, larger mRNA that could potentially encode GAR synthetase or GAR transformylase activities.

Evolution of the AIR Synthetase Sequence

Amino acid alignments placed the bacterial and plant AIR synthetase sequences closer together than the plant and most animal sequences. Analysis of nonsynonymous nucleotide substitutions and an AIR synthetase gene tree based on D_n, although consistent with the amino acid comparison, gave a more complex view. The fraction of D_n is lowest for the Ath/Eco and Ath/Bsu comparison (0.38 and 0.44, respectively). The divergence among the *Arabidopsis* sequence, the two vertebrate sequences, and the two yeast sequences was not much greater (D_n = 0.44–0.48) than the divergence observed among the plant and bacterial sequences. Furthermore, the topography of the gene tree showed only modest resolution among the AIR synthetase sequences from different kingdoms. The plant sequence branched at a position intermediate between the other eukaryotic sequences and the bacterial sequences. Any proposed common ancestor to the bacterial and plant sequence could not have diverged significantly down an independent pathway from the other AIR synthetase sequences. It has been more than a billion years since the AIR synthetase sequences from these kingdoms diverged from a common ancestral encoding sequence. Base substitution rates cannot be expected to be completely linear over these time scales. Additionally, convergent sequence evolution of codons encoding isofunctional amino acids cannot be eliminated as a factor and may have contributed to some of the similarity between the plant and bacterial sequences, which is evident in the topology of the tree.

One might speculate that the monofunctional character of the plant enzyme was an important selective force keeping the plant enzyme more similar to the bacterial enzyme throughout the extensive time since these sequences diverged from a common ancestor. It is interesting to consider that the topographical order of AIR synthetase sequences (Fig. 6B) from bacterial, to plant, to yeast, and then to animal AIR synthetase sequences is concordant with the increased complexity of the enzyme: from monofunctional enzymes in bacteria and plants, to bifunctional enzymes in yeast, and then to trifunctional enzymes in animals. Alternatively, the similarity of the plant and bacterial sequences might reflect an endosymbiotic origin for the plant sequences. The gene would have been transferred at a later date from a protomitochondrial or protochloroplastic genome into the nuclear genome.

Regulation of *PUR5* Gene Expression

The regulation of purine biosynthesis in plants may yield some interesting data regarding other growth-regulated enzymes. Analysis of enzyme induction during nodulation has demonstrated that the enzymes of purine biosynthesis are induced during nitrogen fixation (Atkins, 1991). The mechanisms underlying enzyme induction can be studied using genes isolated from *Arabidopsis* or other organisms. Several significant control points of purine biosynthesis exist in bacteria, primarily at the *purF* and *purA* steps. Using the genes isolated in this study as probes, we can begin a study of the regulation of the purine pathway in leguminous and nonleguminous plants.

We have demonstrated by sequence analysis of a genomic *PUR5* clone the presence of Inr elements near the transcription initiation site. Such elements are known to function independently of a TATA box (Adams and Timberlake, 1990; Pugh and Tjian, 1990), and, indeed, in this gene no TATA box sequences were found within 50 bp of the transcription start site. Regulation of purine nucleotide biosynthetic gene expression has been studied more extensively in other systems. Recent work by Glesne and coworkers (1991) indicates that IMP dehydrogenase mRNA levels are influenced by the intracellular level of guanine nucleotides. This mode of regulation does not appear to be transcriptional; rather, IMP dehydrogenase gene expression is regulated by posttranscriptional nuclear events that are responsive to guanine nucleotide levels. It appears that purine nucleotide biosynthesis may be regulated by a variety of mechanisms, transcriptional and posttranscriptional. Further analysis of the regulatory elements of this gene and additional genes encoding enzymes involved in purine biosynthesis in plants will help to delineate the promoter sequences that are involved in regulating gene expression of this pathway in plants and the mechanisms that are required for biosynthesis of DNA and RNA precursors.

The sequences contained within the 3′-flanking region are likely to contain important determinants for polyadenylation. Although regions within the 3′ end resemble the AAUAAA signal required for efficient polyadenylation in plants, no exact match to this consensus sequence was revealed in our analysis of the *PUR5* 3′-untranslated region. Other workers have also reported that polyadenylation in plants is not as precise as in animal systems and that significant deviation from the consensus elements is not unusual (Shirley et al., 1990). A region 13 to 58 bases upstream of the putative polyadenylation signals is extremely rich in uracil residues (31 of 45). A recent study by MacDonald et al. (1991) demonstrated that the polyadenylation signals for the T-DNA-encoded octopine synthase gene consist of several regions that are high in uracil content. These regions are hypothesized to activate AAUAAA-like sequences for polyadenylation. Such sequences may be shown to be significant for efficient polyadenylation in plants as well.

Another interesting feature of the *PUR5* gene is the presence of a large number of small introns within this gene structure. A genomic clone has been isolated for another gene within this biosynthetic pathway, that encoding SAICAR synthetase (*purC, PUR7* for *Arabidopsis*; J.F. Senecoff and R.B. Meagher, unpublished data). Polymerase chain reaction analysis has indicated the presence of numerous intron sequences within this gene as well (data not shown). Further study will help to elucidate whether intron sequences are important in the regulation of this pathway in plants.

ACKNOWLEDGMENTS

We gratefully acknowledge the generosity of Dr. John M. Smith for his contribution of bacterial mutant strains to this study. We thank Drs. John Mulligan and Ron Davis of Stanford University for the *Arabidopsis* genomic library. We thank Dr. Michael Weise for support provided in using the sequence analysis software. We thank Dr. Lawrence Shimkets, Department of Microbiology, University of Georgia, for critical reading of this manuscript and helpful comments.

We also thank the reviewers of this manuscript for many helpful suggestions.

Received October 2, 1992; accepted February 17, 1993.

LITERATURE CITED

Adams TH, Timberlake WE (1990) Upstream elements repress premature expression of an *Aspergillus* developmental regulatory gene. Mol Cell Biol **10:** 4912–4919

Aimi J, Qiu H, William J, Zalkin H, Dixon JE (1990) *De novo* purine nucleotide biosynthesis: cloning of human and avian cDNAs encoding the trifunctional glycinamide ribonucleotide synthetase-aminoimidazole ribonucleotide synthetase-glycinamide ribonucleotide transformylase by functional complementation in *E. coli*. Nucleic Acids Res **18:** 6665–6672

Aruffo A, Seed B (1987) Molecular cloning of a CD28 cDNA by high-efficiency cos cell expression system. Proc Natl Acad Sci USA **84:** 8573–8577

Atkins C (1991) Ammonia assimilation and export of nitrogen from the legume nodule. *In* M Dilworth, A Glenn, eds, Biology and Biochemistry of Nitrogen Fixation. Elsevier Science Publishers BV, Amsterdam, The Netherlands, pp 293–319

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (1987) Current Protocols in Molecular Biology. Greene Publishing Associates and Wiley-Interscience, New York

Barankiewicz J, Paszkowski J (1980) Purine metabolism in mesophyll protoplasts of tobacco (*Nicotiana tabacum*) leaves. Biochem J **186:** 343–350

Barton JW, Bleskan J, Patterson D (1991) Isolation of a human cDNA encoding amidophosphoribosyltransferase and functional complementation of a CHO Ade-A mutant deficient in this activity. Somat Cell Mol Genet **17:** 311–322

Berry-Lowe SL, Schmidt GW (1991) Chloroplast protein transport. *In* L Bogorad, IK Vasil, eds, Cell Culture and Somatic Cell Genetics of Plants. Academic Press, New York, pp 257–302

Bonner J, Varner JE (1976) Plant Biochemistry. Academic Press, New York, pp 326–327

Burch LR, Stuchbury T (1986) Metabolism of purine nucleotides in the tomato plant. Phytochemistry **25:** 2445–2449

Chen Z, Dixon JE, Zalkin H (1990) Cloning of a chicken liver cDNA encoding 5-aminoimidazole ribonucleotide carboxylase and 5-aminoimidazole-4-N-succinocarboxamide ribonucleotide synthetase by functional complementation of *Escherichia coli pur* mutants. Proc Natl Acad Sci USA **87:** 3097–3101

Cronquist A (1981) An Integrated System of Classification of Flowering Plants. Columbia University Press, New York

Daubner SC, Schrimsher JL, Schendel FJ, Young M, Henikoff S, Patterson D, Stubbe J, Benkovic SJ (1985) A multifunctional protein possessing glycinamide ribonucleotide synthetase, glycinamide ribonucleotide tranformylase, and aminoimidazole ribonucleotide synthetase activities in *de novo* purine biosynthesis. Biochemistry **24:** 7059–7062

Devereaux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res **12:** 387–395

Dower WJ, Miller JF, Ragsdale CW (1988) High efficiency transformation of *E. coli* by high voltage electroporation. Nucleic Acids Res **15:** 6127–6145

Ebbole DJ, Zalkin H (1987) Cloning and characterization of a 12-gene cluster from *Bacillus subtilis* encoding nine enzymes for *de novo* purine nucleotide synthesis. J Biol Chem **262:** 8274–8287

Glesne DA, Collart FR, Huberman E (1991) Regulation of IMP dehydrogenase gene expression by its end products, guanine nucleotides. Mol Cell Biol **11:** 5417–5425

Hall TC, Ma Y, Buchbinder BU, Pyne JW, Sun SM, Bliss FA (1978) Messenger RNA for G1 protein of French bean seeds: cell-free translation and product characterization. Proc Natl Acad Sci USA **75:** 3196–3200

He B, Shiau A, Choi KY, Zalkin H, Smith JM (1990) Genes of the *Escherichia coli pur* regulon are negatively controlled by a repressor-operator interaction. J Bacteriol **172:** 4555–4562

Henikoff S (1986) The *Saccharomyces cerevisiae* ADE5, 7 protein is homologous to overlapping *Drosophila melanogaster* Gart polypeptides. J Mol Biol **190:** 519–528

Henikoff S (1987) Multifunctional polypeptides for purine *de novo* synthesis. Bioessays **6:** 8–13

Henikoff S, Eghtedarzadeh MK (1987) Conserved arrangement of nested genes at the *Drosophila gart* locus. Genetics **117:** 711–725

Henikoff S, Keene MA, Sloan JS, Bleskan J, Hards R, Patterson, D (1986) Multiple purine pathway enzyme activities are encoded at a single genetic locus in *Drosophila*. Proc Natl Acad Sci USA **83:** 720–724

Henikoff S, Tatchell K, Hall BD, Nasmyth KA (1981) Isolation of a gene from *Drosophila* by complementation in yeast. Nature **289:** 33–37

Izui K, Ishijima S, Yamaguchi Y, Katagiri F, Murata T, Shigesada K, Sugiyama T, Katsuki H (1986) Cloning and sequence analysis of cDNA encoding active phosphoenolpyruvate carboxylase of the C4-pathway form maize. Nucleic Acids Res **14:** 1615–1628

Kornberg A (1980) DNA Replication. WH Freeman, San Francisco, CA

Kornberg A (1982) DNA Replication: Supplement WH Freeman, San Francisco, CA

Li W-H, Luo C-C, Wu C-I (1985) Evolution of DNA sequences. *In* RJ MacIntyre, ed, Molecular Evolutionary Genetics. Plenum Publishing, New York, pp 1–94

MacDonald MH, Mogen BD, Hunt AG (1991) Characterization of the polyadenylation signal from the T-DNA-encoded octopine synthase gene. Nucleic Acids Res **19:** 5575–5581

Margulis L, Schwartz KV (1982) Five Kingdons: An Illustrated Guide to the Phyla of Life on Earth. WH Freeman, San Francisco, CA, pp 338

Meagher RB, Berry-Lowe S, Rice K (1989) Molecular evolution of the small subunit of ribulose bisphosphate carboxylase: nucleotide substitution and gene conversion. Genetics **123:** 845–863

Meyerowitz EM (1987) *Arabidopsis thaliana*. Annu Rev Genet **21:** 93–111

Minet M, Dufour M, Lacroute F (1992) Complementation of *Saccharomyces cerevisiae* auxotrophic mutants by *Arabidopsis thaliana* cDNAs. Plant J **2:** 417–422

Moffatt B, Somerville C (1988) Positive selection for male-sterile mutants of *Arabidopsis* lacking adenine phosphoribosyl transferase activity. Plant Physiol **86:** 1150–1154

Neuhard J, Nygaard P (1986) Purines and pyrimidines. *In* FC Neidhardt, ed, Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology. American Society of Microbiology, Washington, DC, pp 445–473

O'Shea-Greenfield A, Smale ST (1992) Roles of TATA and initiator elements in determining the start site location and direction of RNA polymerase II transcription. J Biol Chem **267:** 1391–1402

Price EE, Murray AW (1969) Purine metabolism in germinating wheat embryos. Biochem J **115:** 129–133

Pugh BF, Tjian R (1990) Mechanism of transcriptional activation by Sp1: evidence for coactivators. Cell **61:** 1187–1197

Saitou N, Nei M (1987) The neighbor joining method: a new method of reconstructing phylogenetic trees. Mol Biol Evol **4:** 406–425

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Schendel FJ, Cheng YS, Otvos JD, Stubbe J (1988) Characterization and chemical properties of phosphoribosylamine, an unstable intermediate in the de novo purine biosynthetic pathway. Biochemistry **27:** 2614–2623

Schild D, Brake AJ, Kierer MC, Young D, Barr PJ (1990) Cloning of three human multifunctional *de novo* purine biosynthetic genes

by functional complementation of yeast mutations. Proc Natl Acad Sci USA **87:** 2916–2920

Schubert KR (1986) Products of biological nitrogen fixation in higher plants: synthesis, transport, and metabolism. Annu Rev Plant Physiol **37:** 539–574

Shirley BW, Ham DP, Senecoff JF, Berry-Lowe SL, Zurfluh LL, Shah DM, Meagher RB (1990) Comparison of the expression of two highly homologous members of the soybean ribulose-1,5-bisphospate carboxylase small subunit gene family. Plant Mol Biol **14:** 909–925

Smale ST, Schmidt MC, Berk AJ, Baltimore D (1990) Transcriptional activation by Sp1 as directed through TATA or initiator: specific requirement for mammalian transcription factor IID. Proc Natl Acad Sci USA **87:** 4509–4513

Smith JM, Daum HA III (1986) Nucleotide sequence of the purM gene encoding 5′-phosphoribosyl-5-aminoimadazole synthetase of *Escherichia coli* K-12. J Biol Chem **261:** 10632–10637

Suzuki T, Takahishi E (1977) Biosynthesis of purine nucleotides and methylated purines in higher plants. Drug Metab Rev **6:** 213–242

Tiedeman AA, DeMarini DJ, Parker J, Smith JM (1990) DNA sequence of the *purC* gene enducing 5′-phosphoribosyl-5-aminoimidazole-4-*N*-succinocarboxamide synthase and organization of the dapA-purC region of *Escherichia coli* K12. J Bacteriol **172:** 6035–6041

Van Camp W, Bowler C, Villarroel R, Tsang EWT, Van Montagu M, Inze D (1990) Characterization of iron superoxide dismutase cDNAs obtained by genetic complementation in *Escherichia coli.* Proc Natl Acad Sci USA **87:** 9903–9907

Wolfe SA, Smith JM (1985) Separate regulation of *purA* and *purB* loci of *Escherichia coli* K-12. J Bacteriol **162:** 822–825

Wolfe SA, Smith JM (1988) Nucleotide sequence analysis of the purA gene encoding adenylosuccinate synthetase of *E. coli* K12. J Biol Chem **263:** 19147–19153

Zenzie-Gregory B, O'Shea-Greenfield A, Smale ST (1992) Similar mechanisms for transcription initiation mediated through a TATA box on an initiator element. J Biol Chem **267:** 2823–2830